

NEPHILA CLAVIPES DRAGLINE SILK: APPROACHES TO A RECOMBINANTLY PRODUCED SILK PROTEIN.

Charlene M. Mello, Steven Arcidiacono, Richard Beckwitt, John Prince, Kris Senecal, and David L. Kaplan

INTRODUCTION & BACKGROUND. Spider silks exhibit an unusual combination of strength and toughness that distinguishes them from other natural and synthetic fibers. Silk proteins perform a key natural function as structural fibers, to absorb impact energy from flying insects without breaking. They dissipate energy over a broad area and balance stiffness, strength and extensibility (1,2). In addition to their unusual mechanical properties and visual lustre, silks also exhibit interesting interference patterns within the electromagnetic spectrum (3), unusual viscometric patterns related to processing (4), and piezoelectric properties (3,5,6). These properties suggest they would be good candidates for high performance fiber and composite applications. However, the spider is not capable of producing sufficient quantities of proteins to enable thorough evaluation of their potential. Consequently, we are pursuing recombinant DNA techniques to clone and express adequate quantities of recombinant spider silk for these studies.

In general, silks are formed in sets of specialized glands by insects in the class Insecta and spiders in the class Arachnida. These proteins contain repetitive crystalline domains that influence higher order conformations. The characteristic antiparallel beta sheet structure of many silks, including spider and silkworm silks, assembles through hydrogen bonding and hydrophobic interactions due to the close packing density of the short side chain amino acids in the polymer chains (7). These fibers are insoluble in dilute acids and dilute alkali, resistant to most proteolytic enzymes (8,9,10) and hydrolyzed by concentrated sulfuric acid (11). Lithium bromide, 9M, has been found to solubilize spider dragline silk without hydrolysis (12).

We have focused on the dragline silk fiber from the spider, *Nephila clavipes*. A pair of major ampullate glands within the abdomen of the spider produce the protein which is spun into the dragline fiber. This system was chosen based on previous studies which demonstrated that the tensile strength of the dragline silk from *N. clavipes* was the highest of all silks tested (13). Aside from the genetics, it is essential to develop a comprehensive understanding of the natural processes required for protein fiber spinning. Experiments are underway to begin to unravel some of the fundamentals which may assist us in mimicking the

fiber spinning capabilities of the spider.

Recent cDNA sequence data from the major ampullate gland of *N. clavipes* suggest that there may be two different proteins comprising the dragline fiber; designated spidroin 1 and 2 (14,15). These cDNA sequences share some similarities, in overall organization and in regions of apparent sequence conservation. When translated into amino acid sequences, each polypeptide consists of highly repeated domains. Both polypeptides have polyalanine (n=5-8) regions integrated into larger repeating units. The highly repetitive domains contain consensus repeats of GAGQGGYGGLGGQ for spidroin 1 and GPGGYGPGQQ for spidroin 2. In addition, the C-terminal portions of the two polypeptides are reported to be very similar (15).

PROTEIN CHARACTERIZATION. Characterization of the fiber and pre-spun protein has provided data on amino acid composition, peptide sequences, and the presence of disulfide linkages (12,16). Silk protein has been purified from the major ampullate gland and its amino acid composition compared to that of the dragline fiber (12). Both samples were very similar and resemble the published amino acid composition of spidroin 1. Negligible amounts of proline were detected in our samples of the dragline fiber and the major ampullate gland. It is difficult to find evidence for the presence of spidroin 2 in the major ampullate gland or the dragline fiber due to its very high percentage (13%) of proline. There has been some evidence for variation in composition from spider to spider and even within the same spider at various time intervals during silking which may account for these differences (16,17). Due to this discrepancy, further molecular characterization of the dragline fiber was pursued. Peptide sequences generated from selective chemical cleavage of the dragline fiber and the purified glandular protein also revealed key sequences encompassing the repetitive domains found in the published sequence of spidroin 1, AAGGAGQGGY and LGSQGAGQ. In addition we identified disulfide linkages in the dragline protein. If the fiber is denatured in guanidine-HCl followed by reduction and alkylation with 4-vinylpyridine, a single band can be visualized on an SDS gel with both silver stain and coomassie blue (Mello, unpublished results). We have been unable to identify spidroin 2 in our protein characterization work of the major ampullate gland or the dragline fiber. However, its absence in our studies is not adequate evidence to confirm its absence in the dragline fiber, therefore, we are continuing to search for evidence of the spidroin 2 protein in the major ampullate gland or the dragline fiber.

GENETICS. Protein sequence data were used to generate genetic probes for screening libraries created from *N. clavipes* major ampullate gland mRNA and genomic DNA. Three

probes were designed, the first would hybridize to a repetitive region of spidroin 1, the second hybridizes to a repeat present in the published sequence for spidroin 2 and the final probe would hybridize to the polyalanine regions of both spidroin 1 and spidroin 2. Extensive screening of many cDNA libraries produced a stable 1.8 kb cDNA clone from the 3'-end of the spidroin 1 gene. This fragment encodes for 570 amino acids, or a 48,450 dalton protein. Studies to express this gene in a variety of host systems are underway. Despite the use a spidroin 2 specific oligonucleotide probe we were unable to detect any clones which resembled the published spidroin 2 sequence.

Initial sequence data from our 1.8 kb cDNA clone revealed a single base insertion in the 3' non-repetitive domain not present in the published spidroin 1 sequence. PCR amplification of *N. clavipes* genomic DNA confirmed this observation. Specific primers to the 3'-end of our spidroin 1 cDNA sequence were used to produced a single 300 bp product (18). A comparison of the sequence of the cloned PCR product with the published sequence of spidroin 1 revealed several minor differences: 4 are single base substitutions, while one is an insertion at base 141 of the PCR sequence, corresponding to the insertion detected in our cDNA sequence. The consequences of this insertion are more apparent when the DNA is translated into the amino acid sequence. When the polypeptides coded for by the published sequences of spidroin 1 and 2 are examined, they are extremely similar over a 49 amino acid stretch (see figure 1). However, when the C-terminal portions of the published protein sequences are compared, with the corrected C-terminal sequence of spidroin 1 (Nc-Sp-1PCR), the region of similarity is extended by an additional 54 amino acids. In addition to the increased homology of these sequences, the frame shift changes the stop codon to TAA which is the stop codon used in the published spidroin 2 sequence, and the distance between the polyA tail and the stop codon in both sequences is now very similar. Since both the corrected spidroin 1 and the published spidroin 2 polypeptides share the same conserved C-terminal region, it could be suggested that this would be a feature shared by many or all spider silk genes, or perhaps even by silk genes from other arthropods. In fact, we have obtained a 288 bp fragment of the spidroin 1 gene from *Araneus bicentenarius* which supports this hypothesis (18).

The significance of this highly conserved C-terminal portion of the protein has yet to be defined. One potential explanation is that the non repetitive C-terminal region of these genes is responsible for maintaining the soluble state of silk within the gland (19). An examination of the amino acid composition and predicted secondary structure of this region indicates that it is distinctly different from the bulk of the repeating region of either spidroin 1 or spidroin 2.

Conserved C-Terminal Region

Spidroin 1	RLSSAVSNLVATGPTNSAALSSTISNV
Spidroin 2	RVASAVSNLVSSGPTSSAALSSVISNA
Nc-Sp-1PCR	RVSSAVSNLVASGPTNSAALSSTISNV
Spidroin 1	VSQIGASILVFLDVMSSFKLFSRLFLLLSRS.
Spidroin 2	VSQIGASNPGLSGCDVLIQALLEIVSACVTILSSS
Nc-Sp-1PCR	VSQIGASNPGLSGCDVLIQALLEVVSA LIQILGSS
Spidroin 1	
Spidroin 2	SIGQVNYGAASQFAQVVGQSVLSAF.
Nc-Sp-1PCR	SIGQVNYGSAGQATQIVGQSVYQALG.

Figure 1. Conserved C-terminal Region. Translated amino acid sequences of the publishes *N. clavipes* spidroin 1 and spidroin 2 are compared to the corrected sequence (Nc-Sp-1PCR) in this figure. The areas of 100% homology are highlighted in grey. When the corrected spidroin 1 sequence is compared to the published spidroin 2 sequence the similarity is significantly improved. The two sequences are approximately 75% homologous over 90 amino acids.

Finally, in addition to our natural gene approach, a series of synthetic spider silk genes was constructed based on the sequence information obtained from the native system and optimized for codon usage by *E. coli*. For example, a monomeric gene was designed using a repetitive region of spidroin 1. These monomeric genes were multimerized to produce synthetic genes ranging from 100 to 2000 bp. These genes code for highly repetitive proteins as large as 56,000 Da. The monomer and several of the multimers have been subcloned into expression vectors. One of these genes, a tetramer, produced a 14,000 Da synthetic silk protein purified from *E. coli* lysate (Manuscript in preparation). Solution studies are currently underway to investigate the structural features of this recombinant silk protein. We are currently determining the maximum size multimer of spidroin 1 that *E. coli* can stably maintain and express. In the past we have encountered problems with maintenance of large, GC rich, repetitive silk genes in *E. coli*.

REFERENCES

1. Gosline, J. M., M. E. DeMont and M. W. Denny. 1986. The structure and properties of spider silk. *Endeavour* 10(1): 37-43.
2. Gosline, J. M., M. W. Denny and M. E. DeMont. 1984. Spider Silk as rubber. *Nature* 309: 551-552.
3. Craig, C. L. and G. D. Bernard. 1989. Insect attraction to ultraviolet-reflecting spider webs and web decorations. *Ecology* 71(2): 616-623.
4. Mogoshi, J., Y. Magoshi and S. Nakamura. 1985. Crystallization, liquid crystal, and fiber formation of silk fibroin. *J. Appl. Poly. Sci.* 41: 187-204.
5. Ando, Y., R. Okano, K. Nishida, S. Miyata and E. Fukada. 1980. Piezoelectric and related properties of hydrated silk fibroin. *Reports on Prog. in Polymer Physics in Japan*. 23: 775-778.
6. Fukada, E. 1956. On the piezoelectric effect of silk fibers. *J. Phys. Soc. Japan* 12: 1301.
7. Lucas, F., J. T. B. Shaw and S. G. Smith. 1960. Comparative studies of fibroins I. The amino acid composition of various fibroin and its significance in relation to their crystal structure and taxonomy. *J. Mol. Biol.* 2: 339-349.
8. Tillinghast, E.K. 1984. The chemical fraction of the orb web of *Argiope* spiders. *Insect Biochem.* 14(1): 115-120.
9. Tillinghast, E. K. and E. J. Kavanagh. 1977. The alkaline proteases of *Argiope* and their possible role in web digestion. *J. Exp. Zool.* 202: 213-222.
10. Townley M. A. and E. K. Tillinghast. 1988. Orb web recycling in *Araneus cavaticus* (Araneae, Araneidae) with an emphasis on the adhesive spiral component, gabamide. *J. Arachnol.* 16: 303-319
11. Livengood, C. D. 1990. Silk. In *Polymers-Fibers and Textiles A Compendium*, 789-797 (J. I. Kroschwitz Ed.) Encyclopedia Reprint Series (John Wiley and Sons, New York).
12. Mello, C. M. , B. Yeung, K. Senecal, P. Vouros, and D. L. Kaplan. 1994. Initial characterization of *Nephila clavipes* dragline protein. In, *Silk Polymers, Material Science and Biotechnology*; D. L. Kaplan, W. W. Adams, B. Farmer, and C. Viney Vol. 544 pp.67-79.
13. Zemlin, J. C. 1968. A study of the mechanical behavior of spider silks. Technical Report 69-29-CM (AD 684333), U. S. Army Natick Laboratories, Natick Massachusetts, USA.
14. Xu, M. and R. V. Lewis. 1990. Structure of a protein superfiber: Spider dragline silk. *Proc. Natl. Acad. Sci. USA.* 87: 7120-7124.
15. Himman, M. B. and R. V. Lewis. 1992. Isolation of a clone encoding a second dragline

- silk fibroin. *J. Biol. Chem.* 267: 19320-19324.
16. Lombardi, S. J. and D. L. Kaplan. **1990**. The amino acid composition of major ampullate gland silk (dragline) of *Nephila clavipes* (Araneae, Tetragnathidae). *J. Arachnol.* 18: 297-306.
 17. Work, R. W. and C. T. Young. **1987**. The amino acid compositions of major and minor ampullate silks of certain orb-web building spiders (Araneae, Araneidae). *J. Arachnol.* 15: 65-80.
 18. Beckwitt, R. and S. Arcidiacono. **1994**. Sequence conservation in the c-terminal region of spider silk proteins (spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *J. Biol. Chem.* (in press).
 19. Kerkam K., C. viney, D. Kaplan and S. Lombardi. **1991**. Liquid crystallinity of natural silk secretions. *Nature* 349: 596-598.